TMA4240 Statistikk
2014

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Øving nummer 9, blokk II

Oppgave 1

På ein av vegane inn til Trondheim er UP interessert i å måle effekten av ei holdingskampanje der målet var å få folk til å redusere farten på ei bestemt vegstrekning. På ein dag blei farten på 12 bilar målt. Vi skal gå utifrå at desse målingane er uavhengige og normalfordelte variable med forventing μ og standardavvik σ. Dei tolv observerte fartsmålingane er gitt nedanfor.

x_i:	75	61	85	65	69	82	70	67	62	93	77	74

Det oppgis at $\sum_{i=1}^{12} x_i = 880$ og $\sum_{i=1}^{12}(x_i - \overline{x})^2 = 1034.7$.

a) Forklar kva parameteren μ betyr i denne samanhengen. Skriv opp rimelege estimatorar for μ og σ i denne situasjonen. Kva blir estimata?

b) Forklar kva som meinast med *type 1 feil* når vi utfører ein hypotesetest.

Frå tidlegare undersøkingar har ein at gjennomsnittsfarten (av veldig mange bilar) var på 77 km/t. Tyder resultata frå desse målingane på at forventa fartsnivå på strekninga er lågare enn $77 km/t$? Formuler spørsmålet som ein hypotesetest, gjennomfør testinga og gje konklusjonen. Bruk 5% signifikansnivå.

I punkt **c**) kan du gå utifrå at $\sigma = 10 km/t$.

c) Forklar kva vi meiner med *type 2 feil* og kva som er samanhengen mellom denne og *styrken* til ein test. Gå utifrå at forventa fart til bilane er gått ned til $74 km/t$. Finn sannsynet for at vi i testen i **b**) vil påstå at forventa fart til bilane er blitt lågare enn $77 km/t$.

Finn deretter ut kor manga bilar vi må måle farten til for å få ein test som har styrke minst 0.90 når forventa fart $\mu = 74 km/t$. Signifikansnivået skal framleis vere 5%.

Oppgave 2

Bremselengde for bil med to ulike dekktyper skal undersøkes. En bremseprøve utføres ved at man begynner å bremse når bilen kjører i $80 km/t$ og bremselengde måles. For dekktype 1 utføres n slike prøver. La $X_1, X_2, \ldots, X_n$ betegne bremselengdene målt i disse prøvene. Helt tilsvarende utføres m bremseprøver for dekktype 2. La $Y_1, Y_2, \ldots, Y_m$ betegne bremselengdene målt her.

Anta at $X_1, X_2, \ldots, X_n, Y_1, Y_2, \ldots, Y_m$ alle er uavhengige og normalfordelt. Anta videre at X_1, X_2, …, X_n har ukjent forventningsverdi μ_1, at Y_1, $Y_2, \ldots, Y_m$ har ukjent forventningsverdi μ_2 og at X_1, $X_2 \ldots, X_n$, Y_1, $Y_2, \ldots, Y_m$ alle har samme kjente varians σ_0^2.

Utled et $(1-\alpha)\cdot 100\%$ konfidensintervall for differansen $\mu_1 - \mu_2$.

Regn også ut intervallet numerisk når $\alpha = 0.05$, $n = m = 10$, $\sigma_0^2 = 2^2$ og observerte bremselengder er som gitt i følgende tabell

i	1	2	3	4	5	6	7	8	9	10
x_i	33.0	30.8	28.0	28.7	28.9	26.6	27.9	28.9	27.8	27.4
y_i	23.4	25.3	25.0	28.9	26.7	25.9	24.4	26.8	28.8	25.5

Det oppgis at $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 28.80$ og $\bar{y} = \frac{1}{m}\sum_{i=1}^{m} y_i = 26.07$.

Oppgave 3

Miljøkonsulenten i en kommune ønsker å undersøke den ukjente pH-verdien i et vann. Betegn den sanne pH-verdien for μ. Konsulenten har tilgjengelig to målemetoder. Metode I er rask, men måleresultatene er beheftet med betydelig måleusikkerhet. Metode II er mye mer tidkrevende, men gir mer nøyaktige målinger. Begge målemetodene er velbrukte og variansen i målingene er derfor kjent. Miljøkonsulenten velger å gjøre en observasjon med hver metode. La X betegne observasjonen ved bruk av metode I og Y observasjonen ved metode II. Vi antar at X og Y uavhengige og normalfordelt med

$$\mathrm{E}(X) = \mu \ \ , \ \ \mathrm{Var}(X) = \sigma_0^2 \ \ , \ \ \mathrm{E}(Y) = \mu \ \ , \ \ \mathrm{Var}(Y) = \tau_0^2$$

der σ_0^2 og τ_0^2 er kjente størrelser.

Det oppgis at en forventningsrett estimator (som forøvrig også er sannsynlighetsmaksimeringsestimator) for μ i denne situasjonen er

$$\widehat{\mu} = \frac{\tau_0^2 X + \sigma_0^2 Y}{\tau_0^2 + \sigma_0^2}.$$

Ta utgangspunkt i estimatoren $\widehat{\mu}$ og utled et $(1-\alpha)100\%$ konfidensintervall for μ.

Oppgave 4

I forkant av et stortingsvalg blir det gjennomført en meningsmåling der et representativt utvalg av velgerne blir spurt om de ønsker et regjeringsskifte eller ikke. Anta at andelen av velgerne som ønsker et skifte er p, og la X være antall personer blant n spurte som svarer JA på spørsmålet "Ønsker du et regjeringsskifte ved høstens valg?".

a) Under hvilke antagelser vil X her være binomisk fordelt? Du må relatere antagelsene til situasjonen som er beskrevet i oppgaveteksten.

Anta i resten av dette punktet at andelen av velgerne som ønsker et regjeringsskifte, er $p = 0.7$, og at $n = 20$ personer blir spurt. Bruk at X er binomisk fordelt.

Hva er sannsynligheten for at 18 eller flere av de 20 spurte svarer JA på spørsmålet om regjeringsskifte?

Hva er sannsynligheten for at flere enn 10, men færre enn 15, av de 20 sier JA?

Anta at to aviser på en bestemt dag presenterer resultater fra to meningsmålinger, gjennomført av hvert sitt meningsmålingsinstitutt, Byrå A og Byrå B. La n_1 være antall spurte og X_1 antall som svarer JA i målingen fra Byrå A, og n_2 og X_2 tilsvarende størrelser for Byrå B. Vi antar at X_1 er binomisk fordelt med parametre n_1 og p, og X_2 er binomisk fordelt med parametre n_2 og p, og at X_1 og X_2 er uavhengige.

Vi ønsker å estimere p ved å kombinere resultatene fra de to målingene. To aktuelle estimatorer er

$$\begin{aligned} \hat{P} &= \frac{1}{2}\left(\frac{X_1}{n_1} + \frac{X_2}{n_2}\right) \quad \text{og} \\ P^* &= \frac{X_1 + X_2}{n_1 + n_2}. \end{aligned}$$

b) Finn forventning og varians til hver av de to estimatorene $\hat{P}$ og P^*.

Dersom $n_1 = 500$ og $n_2 = 1000$, hvilken estimator vil du da velge? Begrunn svaret.

Anta nå at $n_1 = n_2 = n$, slik at X_1 og X_2 er uavhengige og binomisk fordelte, med *samme* parametre p og n. Dette medfører at

$$\hat{P} = P^* = \frac{X_1 + X_2}{2n}.$$

Utled et tilnærmet 95% konfidensintervall for p ved å bruke at fordelingen til

$$\frac{\hat{P} - p}{\sqrt{\frac{1}{2n}\hat{P}(1 - \hat{P})}}$$

er tilnærmet standard normalfordelt.

Et tredje meningsmålingsinstitutt, Byrå C, har annonsert at de snart kommer med resultater fra en tilsvarende måling med n_3 spurte. La X_3 være antall som svarer JA på spørsmålet om regjeringsskifte i målingen fra Byrå C, og anta at X_3 er uavhengig av X_1 og X_2. Vi vil nå bruke resultatene fra Byrå A og Byrå B til å predikere hvor mange som svarer JA i den nye målingen. Vi antar i resten av oppgaven at $n_1 = n_2 = n_3 = n = 1000$, og at observerte verdier for X_1 og X_2 er $x_1 = 645$ og $x_2 = 692$.

c) La $Y = X_3 - n\hat{P}$, der $\hat{P} = \dfrac{X_1 + X_2}{2n}$.

Begrunn at det i vår situasjon er rimelig å anta at Y er tilnærmet normalfordelt, og vis at variansen til Y er $\frac{3}{2}np(1 - p)$.

Bruk dette til å utlede et tilnærmet 95% prediksjonsintervall for antallet spurte som i målingen fra Byrå C svarer JA på spørsmålet om regjeringsskifte.

Bestem også intervallet numerisk basert på de observerte verdiene.

Fasit

1. a) $\overline{x} = 73.3, s = 9.7$ **b**) $H_0 : \mu = 77$ mot $H_1 : \mu < 77$, Forkaster ikke H_0 **c**) Må måle farten på 96 bilar eller fleir

2. $[0.977, 4.483]$

4. **a**) $0.035, 0.536$ **b**) $\mathrm{E}[\widehat{P}] = p$, $\mathrm{Var}[\widehat{P}] = \frac{1}{4}(\frac{1}{n_1} + \frac{1}{n_2})p(1-p)$, $\mathrm{E}[P^*] = p$, $\mathrm{Var}[P^*] = \frac{1}{n_1+n_2}p(1-p)$
c) $[633, 704]$